

Collecting, Processing and Maintaining IOI Statistics

Eduard KALINICENKO, Mārtiņš OPMANIS¹

¹*Institute of Mathematics and Computer Science, University of Latvia
29 Raina Boulevard, Riga, LV-1459, Latvia
e-mail: eduardische@gmail.com, martins.opmanis@lumii.lv*

Abstract. In this paper, we describe the whole process of creation, launching and maintaining the IOI Statistics website project. Special attention is paid to the data acquisition and correctness problems. Our experience may be useful for other International Science Olympiads still not having their statistics portals.

Keywords. IOI, statistics, data collection, data maintaining.

1. Introduction

International Olympiad in Informatics (*IOI*, *IOI-WEB*) is the annual International Science Olympiad for high school students. Since the first IOI held in 1989, a certain amount of data involving the competition is accumulated. However, due to a quite different quality of data for different years as well for different aspects of IOI, the data on its own is not useful.

Usually, we need the data to answer specific questions, like the following ones:

- Question 1. What are the results of the competition for the given year?
- Question 2. In what years did the given country participate?
- Question 3. How many medals overall has a certain country achieved?
- Question 4. What results did the particular person achieve?

As far as International Science Olympiads are concerned, Question 1 is answerable as publishing the results usually takes place. Sadly, however, publishing data about the event often ends here. About further questions, to the best of our knowledge, before the work on this project began, only International Olympiad in Mathematics (*IMO*, *IMO-WEB*) had a centralized system that could answer all these Questions. Most of the other efforts in collecting and maintaining the statistics were not made by the organizers. Most notably, Waldemar Gorzkowski and  Adam Tichy-Racs compiled the list of winners for International Physics Olympiad (*IPhO*), but it was done in a separate PDF file, which is not very customizable and exportable and included only medalists.

For International Olympiad in Informatics, the situation was even grimmer. Official IOI website (IOI-WEB) only collected individual IOI results, and often only for the medalists – the issue we will cover more in-depth in Chapter 4. There also existed two parallel projects – the one by Tom Verhoeff (TUE-NL) and another one by Stanislaw Waligorski (EDU-PL). Efforts to bring the results under similar format and process it to display results per country (EDU-PL) were made, but both projects were out-of-date and abandoned. Even the official website (IOI-WEB), despite several renovation attempts, often relied on official individual IOI websites, which have a tendency to go offline in some cases.

The best resource at the time was SnarkNews project managed by Oleg Hristenko (SnarkNews). However, despite being able to answer Questions 3 and 4 (on medalists only), for old IOIs it only had a list of names and countries of all medalists in that year, which was far from ideal. Furthermore, to the best of our knowledge, it did not appear to solve the problems about data aggregation we will touch upon in Chapter 2.

With this in mind, the goal of the project was to collect as many data as possible and aggregate them in a unified format. Further, the goal was extended to create and implement an IMO-like system to be able to explore the data collected. This paper tells the story of this project from its inception to becoming part of the IOI infrastructure. We hope that the lessons we learned could be useful to other International and Regional Science Olympiads looking how to overcome the same problems we tackled.

2. Data Collection

The work on this project began soon after IOI 2011. There was an interest in seeing the cutoffs for past IOIs to observe how they have changed percentile over the years. It was discovered that collecting the scores necessary for each kind of medal was not a trivial task, but after some time the necessary data was gathered (IOI-EDU).

At this time, the scale of the problem IOI had with statistics was observed, as some of the required data was only contained in one of the above-mentioned projects (EDU-PL), which in 2011 was already offline and was only available through the Wayback Machine (Archive). Despite the initial goal to collect only cutoff scores, it was quickly realized that in fact we have collected names, countries and at least total scores of every single IOI medalist. This meant that at least for some sense (in this case, IOI medalists) we can have some basic information, which led to the inception of this project. However, to illustrate the incompleteness of the data available, we can look at Fig. 1 and see that currently only 61% of IOIs have full per-task scores of all contestants available.

Again, limiting ourselves to IOI medalists only, it is trivial to process the raw data in such a way which could help us easily answer Questions 2 and 3. The biggest challenge with aggregating data, perhaps surprisingly, lies in Question 4. It has a couple of problems in its core and none of the resources available at the time appeared to have tackled it. First, the mapping function from a person to a name is not injective. So we cannot identify a person only by the name. Even more, as was discovered during the

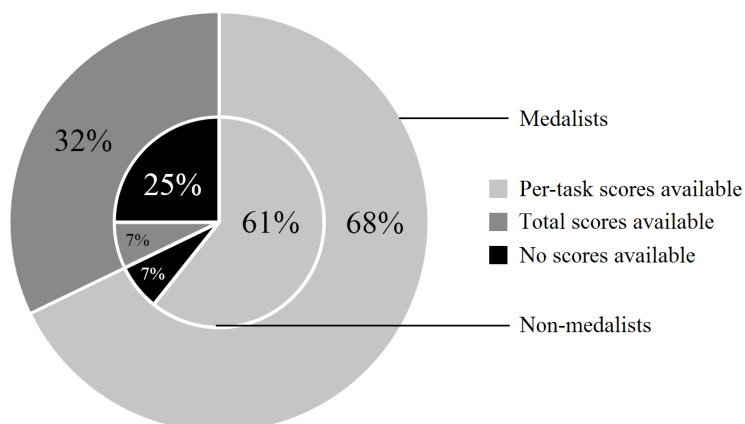


Fig. 1. Availability of contestants' scores for IOI 1989–2016.

collection of the data, even a tuple of name and a country is not enough to identify a person – there are two different people named Chen Zheng from China having participated in IOI.

Even bigger problem is the fact that the function mentioned above is not even a function. Putting aside the issue of a person changing a name (which in fact had occurred before – Jeppe Hallgren was known as Jeppe Petersen during IOI 2010), there is a serious issue of using a different spelling for the same person across different years. Most notably, Gennady Korotkevich was sometimes spelled as Henadzi Karatkevich. This whole issue is easily avoided with the help and understanding of the problem by the organizers – for example, the current IOI registration system asks leaders to pick specifically whether any delegation member is already in the registration database. However, until this support was added, this crucial part of information was lost, and so the biggest challenge in the whole data collection process was to recover this information.

In general, it was found that the same person is likely to represent the same country, although of course some exceptions are occurring like Fieke Dekkers (full name: Sophia Antonia Janna Dekkers), who was a member of a delegation from both Netherlands and the United Kingdom or Rob Kolstad representing the USA and the United Kingdom. Even limiting ourselves to contestants, which we are at this stage, Tomasz Czajka has won two gold medals – one for Poland and one for the United Kingdom. Furthermore, it is reasonable to expect that if it is the same person, then he has competed in close to consecutive IOIs – even if not strictly consecutive, a maximum gap of two IOIs is expected. Again, this does not hold if we are considering all delegation members, especially because it is common for former contestants to appear as a delegation member after some gap. However, at this stage, these two assumptions turned out to be a good basis for determining whether two differently spelled or written names are in fact the same person.

It was decided that it is not realistic to expect that decision can be made automatically without human input. However, it is certainly possible to provide a good set of

candidates for a human to look at in an attempt to minimize the human effort required. For the spelling differences mentioned above (which usually were a lot smaller than in the case of Gennady Korotkevich), Levenshtein distance (Levenshtein) between strings was quite a good indicator – if it were below a certain threshold, we would need to look at the pair to determine whether they are in fact the same person. However, it did not help to identify the second common case – if a person has several middle names, then often one set of middle names appeared in one year and a different one in another. In an attempt to tackle this issue, Levenshtein distance algorithm was run between all possible subsets (with the size at least 2) of individual words from the original strings. If any of those distances were small enough, the pair got triggered. In retrospective, while this did not catch all the cases, we feel that it generated a good enough accuracy in the original data set released.

Instead of unifying everything to a single format, as much data as we could find was released as a single Excel file (IOI-EDU). This included some tricky bits which were not documented well. For example, at IOI, the host can participate with the additional teams. At present, these teams are not competing officially, and so their representatives simply appear in the overall results at the proper position by the score but without any rank or medal. However, nothing of a sort was mentioned in the regulations earlier on. This led to a couple of precedents. Overall IOI 1989 winner, Teodor Tonchev, represented a second Bulgarian team. However, the official IOI 1989 printed booklet ranks him and declares a winner, so this interpretation was kept for that year. Secondly, in 1996 the second Hungarian team won four bronze medals. They were ranked in all the data we have seen and as per our inquiries, they appear to have received physical medals, so they were kept ranked as well.

To conclude, in this Chapter, we have discussed the challenges encountered regarding collecting and processing the data. The most important issue we have identified here is to avoid having a person's name and surname as the unique identifier. Even without a centralized system organizers should make efforts to record and keep information that can be used to tackle this problem in future.

3. Website Planning

At this point, there were no immediate plans to carry on. The data has been collected, and it has even received sets of corrections, in particular by Ilham Kurnia and Mojca Miklavac. However, creating a convenient interface to host the data collected was the end goal – while the created resource (IOI-EDU) was still much more useful than anything available was at the time, it felt that there would be a huge potential wasted if it did not evolve further. To the best of our knowledge, while some efforts to create an IMO-like website based on the data collected were stated to us, they did not transform into anything concrete. That is the reason why in the spring of 2012 it was decided that the matter should be put in our hands and as such the plans for devising the website began to form.

The most crucial thing to get right was the database design. It is tempting to incorporate a lot of redundant information in the tables so that obtaining the data necessary is as easy as possible. However, that could lead to huge problems with maintaining the tables. As such, the effort was spent into carefully designing the database format to minimize the amount of redundant data kept.

However, there are some notable exceptions to that in the database design. For example, the rank and the medal contestant earned is kept in the competition record. Technically this information is redundant since it can be obtained by analyzing others results. Such decision was taken, firstly, to speed up the computation, as it eliminates the need to query all that olympiad's results if we are just interested in a single person; and secondly, to ease the handling of unranked teams, as now we can explicitly set a contestant to be unranked regardless of his position. The other example would be keeping the score of a contestant in his records directly. This is not always a redundant information since we do not always have full results (including the per-task scores). However, even if we had, there could have been some additional problems – at IOI 2000 to avoid zeroes at the end of the competition extra 50 points were awarded per day for simply turning up, so in that case even if we had full results, the contestant's total score would not simply be the sum.

The next big decision is about the cooperation between the database and the web server. It was thought that the database design was much less likely to change over time than the web server, and as such it was decided to write complicated queries which take care of everything, including composing and sorting the tables requested. This means that it is incredibly simple to write a website, as shown in Fig. 2, which queries the data-

Participant: Gennady K... × +
<http://localhost/ioires/participants?id=804>

Year	Country	Tasks												Score		Rank		Medal				
		Abs.	Rel.	Abs.	Rel.																	
2006	Belarus	W	100	P	59	F	73	M	100	J	26	G	21	379	63.17%	26	91.13%	Silver				
2007	Belarus	A	80	F	55	S	80	M	100	P	56	T	30	401	66.83%	20	93.33%	Gold				
2008	Belarus	P	100	F	45	I	22	T	100	G	100	P	70	437	72.83%	7	97.88%	Gold				
2009	Belarus	A	43	H	100	P	100	R	100	G	100	M	100	R	100	S	100	743	92.88%	1	100.00%	Gold
2010	Belarus	Q	100	L	100	H	95	C	100	S	100	M	83	T	100	M	100	778	94.88%	1	100.00%	Gold
2011	Belarus	G	100	R	100	R	100	C	100	E	100	P	100	600	100.00%	1	100.00%	Gold				

Participant: Atamurad ... × +
<http://localhost/ioires/participants?id=296>

Year	Country	Tasks												Score		Rank		Medal
		Abs.	Rel.	Abs.	Rel.													
2004	Turkmenistan	A		H		P		E		P		F						
2005	Turkmenistan	G	25	M	0	S	100	B	54	R	70	R	35	284	47.33%	130	53.26%	Bronze
2006	Turkmenistan	W	100	P	100	F	49	M	80	J	35	G	25	389	64.83%	20	93.26%	Gold

Fig. 2. Simple website to query the contestants' results from the database.

base, once the queries are known. The website itself looks quite different nowadays, but this table is remarkably similar to the one used today.

Initially, there was a third goal conceived, which was to unite all International Science Olympiads to be able to view a person's results across all competitions. Due to lack of workforce, we decided to focus specifically on IOI instead. However, the database was designed with more than one science olympiad in mind and as such the database was split into two schemes – the one with IOI-specific data, like competition records, and the one with the common information (like information about people and countries). As a result, it would be easy for some other olympiad to use the information from the common information scheme, allowing people and countries to share information across olympiads.

The website itself was coded in PHP, however as mentioned previously, most of it is just the code for displaying the information in a pretty way on the website. The biggest part of the website devoted to this part is request processing, where we need to combine all information obtained from forms to a single SQL query, resize photos, etc. However, it would not take a ridiculous amount of effort to replace the web server with another implementation of it. As for the actual design and structure of the website, it was deliberately designed similarly to IMO's statistics. However, certain aspects were improved. For example, the country delegation page with photos is only easily accessible while IMO has not finished yet, and we made sure that the accessibility of that page is improved in the IOI statistics.

Finally, what is important to consider but easy to miss is the URL lifetime. Ideally, you would like to make sure that the link created a long time ago would still work in the future and would not become dead. This means that it is desirable to create a good set of URLs at launch so that they are as human-friendly as possible. For example, compare http://www.imo-official.org/country_individual_r.aspx?code=LVA and <http://stats.ioinformatics.org/results/LVA>, where both of these links lead to the same type of page. The major changes are also mostly backward compatible. For example, the link which could have been made in 2012 (IOI-GDR) still works, despite the server move and a country code change to comply with ISO standards. The only dead links resulted after merging the entries of two persons (usually, after an e-mail that two entries which we treat as different people are in fact the same person), and then using the old link would not redirect to the new one, however this will get fixed if we can reconstruct with confidence which entries were merged from backups.

In retrospective, there were also a couple of bad design decisions. One of such is the "Login" button. Its sole purpose is to allow the moderator to login into the system to view and act upon submitted requests. However, it sometimes creates confusion when users think that they are required to have an account to submit a change, while in fact "Edit" or "Add" interfaces are available to anyone. Secondly, in the "Edit" interface, there is a "Submit" button after every section (Basic Information, Contact Information, Participation Information and Photographs). The reasoning behind this was a separation of requests by category with the idea that if additional verification is needed for confirming an IOI role, we could accept the changes of contact information separately. It turned out to be a bad decision, as on several occasions people had missed multiple

“Submit” buttons and were pressing the last one, which resulted in the loss of data when users thought they submitted a change while only the photo was submitted. Finally, after submitting a request, one receives a password to revoke it if needed. Sometimes people seem to believe that this is the CAPTCHA-like process for the request and accidentally cancel them instead. Perhaps, revoking requests in a suggested way is not usual, and additional warnings are required to prevent misunderstanding.

To conclude, in this Chapter, we have discussed the challenges encountered while designing a website. We feel that it is crucial to spend a considerable amount of time on planning the back-end, which can be easily overlooked in favor of the front-end. We feel that without a reasonable back-end, maintenance becomes unmanageable, which often leads to project abandonment.

4. Acquiring Data

The biggest challenge was obtaining the perfect (correct and complete) data. We currently have data about 5775 people in the database. However, a lot of data is still missing. Beginning from the first Olympiad at 1989 hosts usually produced a full list of the current IOI participants. Since 2000, this duty is stated in IOI Regulations (IOI-REG ver. 2014 S5.10(4)). Nowadays, publishing such a list is straightforward since all the data is in the IOI Registration system and it is given to the host for organizational purposes. However, now, we have no such official source of information for years 1990, 1993, 1995, 1998, 2002. Of course, now we can just point to the fact that these essential parts of IOI history are not kept in an appropriate manner.

During the initial project (IOI-EDU), only information about contestants was collected. Now, we were suddenly interested in obtaining information on team leaders and other delegation members. We limit ourselves just to “international part” of IOI participants. In the IOI statistics, you cannot find information about representatives of the host country (management, technicians, guides, volunteers, etc.) who help to organize the whole event. Also, from the historical perspective, IOI roles are changing. For example, during the first IOIs team leaders comprised an International Jury not existing anymore as a body. In 1993, there were such positions as “President of the International Olympic Committee” and “President of the Technical Committee”. According to Regulations for that period, a president was nominated by the host country and today seems to correspond to the role of Chairman rather to the position of the President of IOI elected for a three-year term. Taking into account that it is almost impossible to solve all these semantic clashes of the past, we defined the following twelve roles of IOI participants: Contestant, Leader, Deputy Leader, Guest, Invited Guest, Observer, President, Executive Director, Chairman, International Committee (IC) Member, International Scientific Committee (ISC) Member, and International Technical Committee (ITC) Member.

However, even keeping names of roles as close to the current Regulations as possible, there appear unexpected problems. For example, according to the Regulations

current host country has one representative in the IC. But how to proceed if the host is represented by different people in the in-between meeting of IC and at the IOI itself? Are there two representatives of the host or one person (which one?) must be forgotten? Also, in an attempt to avoid expanding the list of roles too much, by “Invited Guests” there are marked different categories of people – task authors invited by the Scientific Committee, invited members of international organizations and associations and others.

We also provided functionality for the user to add some personal information to his profile like Codeforces/TopCoder handles, homepage, e-mail and social network profiles. This was implemented partially due to the anticipation that the website might be used for direct contacts and recruiting purposes by some parties and as such the ability to add contact information was desirable.

While the situation for contestants was not perfect, but still reasonable, the situation for other members of the delegations was close to appalling – even national websites listing all IOI teams for a specific country often omitted team leaders. However, it was anticipated that national delegations at large still have access to that information, so the natural choice would be to allow delegation leaders to provide and correct simply their data in the system. However, that did not quite solve the problem of contact information, and as well as that, it heavily relies on all delegation leaders being responsive to the cause. So it was decided that the editing access would not be limited to a single person for every country.

Instead, we decided to make the editing publicly available without any registration. Compared to the denied approach, we realized that there would also be bad and malicious edits. However, our estimations said that this way we would receive more data; and if we can filter the good requests from the bad ones, we would obtain more information this way, and since this was our primary goal, we decided to go with that. On the downside, every change submitted to the system has to be approved by moderators, which created an additional effort on our side of things. In retrospect, there were 12818 non-spam requests (including rejected malicious ones) at the time of writing out of 1295397 requests received overall, meaning that 99% of the requests we receive is spam. However the vast majority of them never reach our eyes, so despite that, we feel that it was the correct decision.

Furthermore, sometimes it benefited to go through a central point of moderation. For example, some confusion was created with Sweden and IOI 1989. The national website for Swedish Olympiads in Informatics lists IOI team for 1989 (SOI). Similarly, we have received requests to add this team to the system. However, we were in possession of the physical copy of the first IOI booklet having no records about a team from Sweden. After some conversations with Pär Söderhjelm, former Swedish team leader, it was learned that in 1989 there were two similar IOI-like initiatives. The Swedish team picked the “wrong” one (we would be interested get more information about this event!), so they did not participate in IOI 1989, and only joined IOI in 1990.

There were rare cases in the IOI history when teams mentioned in the official booklets did not participate in the competition due to visa issues or other, mostly political, reasons. Such cases of disagreement with official documents were resolved using direct communication with country representatives.

The biggest issue regarding missing data was generated by the rule that IOI had from 2000 – for non-medalists their names and represented countries should not be published (IOI-REG ver. 2014 S5.10(3)) in the results and as a consequence usually are lost from freely available sources of information. We believe that not mentioning names of non-medalists is contrary to Olympic spirit formulated by Pierre de Coubertin “The most important thing in the Olympic Games is not winning but taking part; the essential thing in life is not conquering but fighting well” (Coubertin). Contestants qualified to IOI as a rule passed many national and regional events and are good even if they do not obtain any medal at IOI. Furthermore, it creates a huge gap between the last bronze medal and the first person without the medal. Despite these two people being right next to each other in the standings, this rule made impossible for the second one to highlight his achievement easily. But for this project, it was decided that we would not allow individuals to de-anonymize their results if they desire as this was deemed to be impossible to verify. This concern turned out to be largely irrelevant, as we did not receive any such requests.

In an attempt to at least not to lose these names for the history, in the IOI Regulations since 2014 is stated that current IOI host is obliged to “Produce a full result list containing the final scores of all contestants, which is made available to the OED and ISC, along with the data required to generate those scores;” (IOI-REG ver. 2014 S5.10(4)). On the public side of things, this rule began being pointless from IOI 2010, when the live scoreboard during the contest was introduced. As such, other projects like SnarkNews (SnarkNews) have full results available and so it became a pointless exercise to remove the non-medalists’ names as the full results were already made available previously. In a hilarious display, IOI 2010 website still hosts both versions of results – anonymized and de-anonymized. As such, we have taken a strict stance on maintaining the full results of IOIs from 2010 onwards. Since then the risk of losing essential contest information is diminished.

However, we fully respect this rule while no live scoreboard was available. Most notably, during the data collection stage, we have managed to discover the cached full results of IOI 2004, which presumably were published by organizers by mistake and were taken down soon afterward. Including these data in the Excel file resulted in two e-mails from IOI 2004 contestants asking to remove their name from the results. Their wishes were respected, and this was also the main reason for the decision of not publishing the de-anonymized results of IOI 2004’s non-medalists at all on the present website.

The biggest issue with that rule, however, is the fact that it made it easy for IOI organizers to neglect the data. Since 2000 in IOI Regulations it was specified that three lists should be published after the IOI:

1. The final scores of the medal winning contestants.
2. A list of all participants.
3. A list of all scores which contains no name/country information (IOI-REG, ver. 2014 S5.10(3)).

This rule does not oblige host to keep full scores for all contestants in any form. Moreover, often only the first list (as the most important list of the whole event) was published, meaning that non-medalists scores and all participants’ names were lost even-

tually. We believe this is the main reason why the data we have today is so incomplete, which we believe was not anticipated at the time this rule was passed. Closing discussion about the availability of full results, it should be added that also rare cases of disqualifications are not clearly marked in results. Usually, such contestants appear as the last ones in the full results of the particular IOI together with contestants having no single submission into grading system.

Despite the fact that IOI is an individual competition there are also some cumulative statistics available regarding country participation in IOI: the total number of medals, the total number of participated contestants, years of participation (counting only the years when there were contestants on the team), etc. However, it is important to note that we still keep records of all years even when there was only a single observer in a delegation. Following the spirit of individual competition, countries are never explicitly ranked; there are different ways to sort data, but you will never find rank next to the name of a particular country.

To conclude, in this Chapter, we have discussed the challenges encountered while attempting to collect additional data to obtain the complete picture and why this was necessary in the first place. The takeaway message from this would be to make sure that hosts publish all required information, and then make sure that the information is duplicated in some centralized location – sometimes the last step would be forgotten, and the information would be lost after the host website would go offline. While Web Archive might save this information, it is not a good idea to rely on this.

5. Moderation

After reading the written above, a reader may have an impression that work is already completed and new data may appear only after the current IOI, and there will be no updates in-between IOIs. Concerning IOI statistics, this is far from the reality. There is still plenty of work to perform outside of IOI period. In this Chapter, we list several examples of such work in the decreasing order of importance.

First, the main effort goes into the attempts to fill in gaps in the history by trying to contact persons probably having missing documents. We are confident that it is impossible to get official data from hosts for five IOIs mentioned above. The most realistic way is to contact former team leaders and try to fill gaps on “delegation by delegation” basis. The main obstacle here is that during early IOIs generations are changed and without proper local archives it is almost impossible to restore information. As well, different countries are differently interested in completing information regarding their country. At the moment of writing paper statistics for 44 countries (from 101 or 43.56%) are completed.

Next, we have to deal with major requests concerning the website. These requests usually affect the whole website, so they have to be dealt with carefully. To give an example, in the last season, two of such requests stood out.

For the first one, a former contestant argued that his name must be removed due to “right to be forgotten”. We insist that IOI Statistics is the source of strictly statistical and

historical information, and we do our best to keep it correct and complete as possible. Removing any essential piece of correct IOI information is against our efforts to complete the whole picture of the IOI history. Also described “right to be forgotten” cases influenced just search engines (how easy is to find a particular piece of information) not the availability of information itself.

For the second example, S. Maggiolo suggested adding gender information to contestants’ data on the official IOI statistics website (Maggiolo). Since we rely on authoritative sources of information and no gender information was kept until the creation of the IOI Registration system, we do not believe it’s feasible to collect and verify the gender of individuals. Furthermore, addressing contestants according to their gender itself might be questionable – we are not attaching other personal information like race, religion, food preferences or disabilities; and we believe that some people might object to releasing information on their gender as well. We understand that these statistics might be considered useful to answer some research questions, like whether vegetarians tend to be better programmers, but we are not convinced at this point that the official IOI website is the right place for that. This matter was recently raised with International Committee.

Moving on, we have to deal with processing the usual requests. Commonly, these are about changing/adding photos or updating contact information. If there is a request to change contact information, requests are verified by checking a proposed link and its relevance to the addressee. In unclear cases if contact information is known, before approving, subject to the proposed changes or country representative is contacted. However, we periodically receive requests by e-mail (because these types of requests are not supported by the system). Most commonly, the request is to merge two of the profiles together because in reality it is the same person, which is usually quite easily verified.

Of course, not every request is legitimate and as such we have to struggle with hoaxes and improper requests. Relatively often, there are requests to change information in an appropriate manner. For example, a request to add “contestant of team X” for the year where we are confident that we have complete data (even more if suggested name is like “Anonymous”). Or request to remove all contact information for some participant, which appears to be fake after consultation with the addressed person. Or adding an improper photo like well-known Borat dressed in the green outfit (see Fig. 3). In general, almost all such requests seem to be malicious than just joyful. Finally, there is also an issue of dealing with spam. However, most of the work here is done automatically, and just most sophisticated cases pass through the spam filter.

For readers, it may be interesting to know how the moderation of requests submitted through the system looks from the inside. An example screenshot is shown in Fig. 3.

In the section “Administration” there is a possibility to clear spam, approve or reject a batch of requests according to submitter’s IP address. Note that we can also change the availability of the data for certain IOIs. This comes useful during the IOIs when the results can be entered in the database long before they become official (and they will not “leak” to the public), and we can then publish them with one click as soon as they become official. Section “Country Data” offers a possibility to change color code for the specified country in the main table of countries describing information completeness:

The screenshot shows the IOI Moderator interface. At the top, there is a logo for IOI and the title "International Olympiad in Informatics – Statistics". Below the title are navigation links: "Olympiads", "Countries", "Tasks", "Hall of Fame", "Search", "Add", "Requests", and "Logout".

The "Administration" section includes buttons for "Make IOI 2015 online", "Make IOI 2015 offline", "Clear spam", and a form for "Process pending requests by IP:" with "Grant" and "Reject" buttons.

The "Country Data" section has a dropdown menu for "Make" (set to "Albania") and a "status" dropdown (set to "green"), with an "OK" button.

The "Requests" section contains a table with the following data:


ID	Time created	IP address	Changes	Action
1290898	2016-04-12 13:01:53	x.x.x.x	Changing Mārtiņš Opmanis: Twitter: https://twitter.com/MartinsOpmanis	Grant Hold Reject Spam
1205317	2016-01-26 21:20:15	y.y.y.y	Changing Borat Jurcić Zlobec: IOI 1994 photo: 	Grant Hold Reject Spam

Fig. 3. Moderator interface screenshot (IP addresses are obfuscated).

green – information complete, yellow – information almost complete (usually completed regarding contestants), red – essential part of the information is missed.

All requests must be approved (button “Grant”), rejected (“Reject” or “Spam”) or have the processing postponed (“Hold”). In the given example, both requests are real. The first one would like to add a Twitter profile to person’s requisites and, therefore, will be approved. The second one is obviously non-legitimate so that it would be rejected. It is worth noting that even if the photograph were of the person in question, it would still get rejected as we aim to have more passport-like photographs on the website (although this desire is not expressed clearly), where the head occupies a significant percentage of photo’s vertical space.

To conclude, in this Chapter, we have discussed the challenges encountered while maintaining the website. We feel that the most important issue is not to take accepting any edit requests lightly. While it is understandable that occasionally some false information provided by the community might slip through, if this information is false, one would lose the reputation which is crucial if you want to be a source of credible information.

6. Launch and Steps to Becoming Official

The statistics website was launched during IOI 2012 after a presentation during the IOI Conference. The results of IOI 2012 itself were added about a week after it was finished. The website accumulated an average of 140 unique daily IP visits during first two months after launch. The appeal to obtain more information was also fruitful. In the period until the next IOI we have granted 1233 requests, whether they were submitted

by the community entirely, or we have committed them ourselves based on the raw data submitted by the community.

Meanwhile, in the same period, we have received access to the official IOI website (IOI), which at the time was quite outdated. This meant that we could start to plan on how to move this project under the official domain. Initially, the idea was to incorporate essential information about IOI on the project and replace the original IOI website completely. However, this was later abandoned for cleaning up and maintaining the original website and hosting the statistics portal as a separate subdomain.

During IOI 2013, it was voted by the General Assembly (GA) to sponsor maintenance of the original IOI website and the development of this project from the IOI budget (IOI-2013-GA-MIN). This allowed us to fulfill the goal mentioned above and in April 2014 the website was moved to the place where it currently resides (IOI-STATS). While we were still supplementary to the official results published by the organizers, for IOI 2015 (IOI-2015) they did not do so and instead simply linked to our project, which we consider being a final step in becoming official.

One of the other aspects of becoming official outside of recognition is the ability to access parts of official data. This is best showcased by the process of releasing the list of participants before the IOI occurred. Usually, every year well in advance of IOI there appears a popular thread on Codeforces and TopCoder, where the community would share its knowledge on the upcoming IOI contestants, and that would be summarized, creating an unofficial IOI contestant list. The ability to display participant list was first added to the website for IOI 2013. At that stage, we did not have any access to the official data in the registration system (IOI-RS) and as such, we simply mirrored the thread on Codeforces for that year. After the organizers had released the official list, the website's data was modified to reflect that.

For IOI 2014, we decided to keep it official and did not take any data from the community, as some talks on accessing official data began already. Unfortunately, it did not reach the conclusion in time, and as such we have waited until the official data was released. Strangely enough, initially it got released exclusively on the mobile app for Android. Still, it was mirrored to the database and for some time, it was the only convenient way to view contestant list if you do not have an Android device. Subsequently, we received the necessary data (name, surname, role, and country – note that we always receive information on per-need basis and do not receive unrelated and personal information to the cause, like passport details, for obvious reasons, including the one of privacy) from the official IOI registration system, so we could add some additional delegation members (like guests, who were not displayed in an app). Also, we discovered a discrepancy between the registration data and the data published by the organizers, which was then addressed and fixed.

In the meantime, this issue was discussed within IC to ensure that we could receive the data from the IOI registration system sooner. Additionally, the check mark was added to the system to allow us to publish the participants' photographs as well. Because of these, for IOI 2015 we were able to publish the participant list soon after the registration deadline was over and we were the first resource to do so. We were able to publish 228 photos, where we were given permission to do so. Extra effort was made to avoid ac-

cidental leaks of personal information, where some people uploaded a full passport copy instead of a photo. Even despite the fact that we shrink the photographs to 180 pixels in height, and it would be very unlikely that there would have been any leaks because of that, we still manually cropped those photographs.

Finally, the process of becoming official means that you receive a lot more credibility over time and as a result people start relying on and linking to your information. At the moment of writing this paper, there are over 9000 links to “stats.ioinformatics.org” according to Google search. Some of them are not only simply linking to results or individual statistics, but also refer to actual statistics. For example, the post on Quora (Quora) was attempting to answer the question of the hardest IOI problem based on the average score per contestant available on the website. Of course, not all of the links are of the same level of importance. Moreover, there are incorrect ones. For example, link in the article (UG-RU, 2016) claims that there will be problems from the corresponding IOI where there is currently just statistical information about tasks.

Investigating pages with links, we found one, which may be the example why we cannot take any responsibility for the way how provided statistical information is used. For example, in the forum post (Apricity) photos from the IOI Statistics regarding Chilean contestants are extracted and an attempt to evaluate demographics based on ethnicity was made, which is not something that we approve or imagined that would happen based on the data we released.

To conclude, in this Chapter, we have discussed the process the website took to obtain an official status. Furthermore, some benefits of the official status were provided.

7. Future Work

One of the evergreen tasks is encouraging people to fill in missing parts of information. Just as a reminder for potential submitters – information can be submitted in a few simple steps. First, consult the “People” page of the particular country to find the particular person whose information you would like to update. If such a person already exists there, follow the link under that person’s name and push “Edit” in the top-right corner of the page. After adding/changing appropriate information, click the “Submit” right after changed section. Be aware – there are four sections having separate “Submit” buttons, so if you want to edit information in multiple sections, you would need to submit multiple requests to the system! Only if you cannot find the person in the list of already known people, push “Add” and provide all known information about the particular person. In this case, “Submit” must be pushed just once at the very end. As a backup scenario, there is always a possibility to send all relevant information to us – we will add it by ourselves. The main principle – do not keep valuable information about former IOIs a secret!

As far as the further development of the website is concerned, the current plan is to refactor the web server code completely. When it was created, not many efforts were put into making it maintainable and as the result the code became quite unreadable over time, and it is now quite challenging to add new features without accidentally break-

ing something. After that, it would become feasible to considering open-sourcing this project so that other olympiads, like Regional ones, could use this platform to host their results. As well as that, perhaps some additional functionality could be contributed by the community then, as is currently happening with CMS (CMS), which is the competition system currently used at IOI.

8. Conclusions

In this paper, we have touched upon various aspects and challenges we have encountered in creating and maintaining a centralized place for collecting official IOI data. While it is not a small project and some International and Regional Science Olympiads might not have the resources to tackle this at the moment, we feel that it is still important to consider our experience. While the project itself is hard work, many challenges that made it hard are solvable easily if they would have been considered at the time. So we hope that after reading our experience some Olympiads might pay a bigger attention to the issues discussed here and make it a lot easier to execute the project of this scale.

Acknowledgements

We would like to thank Oleg Oshmyan for all the technical help received during the creation of this project and maintenance of the servers. We also extend our gratitude to Sergey Melnik and Vyacheslavs Kashcheyevs for supporting this project and giving us fresh ideas. We would also like to thank Ilham Kurnia and Mojca Miklavec for providing a massive amount of data in the early stages of this project. Finally, we thank IOI community for providing the financial support for this project.

References

- Apricity. *The Apricity. Why do people still think that Chile doesn't have a mestizo majority?*
<http://www.theapricity.com/forum/showthread.php?184231-Why-do-people-still-think-that-Chile-doesn-t-have-a-mestizo-majority>
 Archive. *Internet Archive: Wayback Machine*. <https://archive.org/web/>
 CMS. *Contest Management System*. <https://cms-dev.github.io/>
 Coubertin, P. *Quotes*. http://www.brainyquote.com/quotes/authors/p/pierre_de_coubertin.html
 IOI-PL. *Results of International Olympiads in Informatics, formerly at*
<http://www.oi.edu.pl/ioires/>
 Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR*, 163(4), 845–848.
 IMO-WEB. *International Mathematics Olympiad*. <http://www.imo-official.org/>
 IOI-2013-GA-MIN. *Minutes of the General Assembly of the 25th International Olympiad in Informatics*.
http://ioinformatics.org/a_d_m/ga/ioi13/GA-minutes-jul2013.pdf
 IOI-2015. *The 27th International Olympiad in Informatics*. <http://ioi2015.kz/>
 IOI-EDU. *International Olympiad in Informatics: The Results*. <http://www.eduardische.com/ioi/main>

- IOI-GDR. *Information regarding the former German Democratic Republic in the statistics of the International Olympiad in Informatics.* <http://ioi.eduardische.com/countries/GDR>
- IOI-REG. *Official Regulations of the International Olympiad in Informatics.* <http://ioinformatics.org/rules/index.shtml>
- IOI-RS. *Registration System of the International Olympiad in Informatics.* <https://ioiregistration.org/>
- IOI-STATS. *Statistics of the International Olympiad in Informatics.* <http://stats.ioinformatics.org>
- IOI-WEB. *International Olympiad in Informatics.* <http://ioinformatics.org/>
- IPhO. *International Physics Olympiad statistics.* <http://ipho.org/statistics.pdf>
- Maggiolo S. (2015). An Update on the Female Presence at the IOI. *Olympiads in Informatics*, 9, 127–137. http://ioinformatics.org/oi/pdf/v9_2015_127_137.pdf
- Quora. Thread “*Specific Problems in International Olympiad in Informatics*” <https://www.quora.com/What-are-the-hardest-problems-from-past-IOIs>
- SnarkNews. *SnarkNews on IOI.* <http://ioi.snarknews.info/>
- SOI. *Swedish Olympiads in Informatics.* <http://www.progolymp.se/1989>
- TUE-NL. *Former website of the International Olympiad in Informatics.* <http://olympiads.win.tue.nl/ioi/>
- UG-RU (2016). Information about the 28th International Olympiad in Informatics. *Uchitelskaya Gazeta*, Number 4. (In Russian. Учительская газета). <http://www.ug.ru/archive/63529>



E. Kalinichenko is in his final year of studies at the University of Cambridge towards BA/MEng in Computer Science. He obtained four medals at IOIs, including a gold medal at the IOI 2011 and had obtained an honorable mention at IMO 2010. Participated in two ACM International Collegiate Programming Competition World Finals – one as a contestant from University of Latvia and one as an on-site coach of University of Cambridge team. He is a member of the Latvian jury for high school programming competitions and has been a deputy leader of Latvian delegation at IOI 2013.



M. Opmanis is a researcher at the Institute of Mathematics and Computer Science of the University of Latvia. He is one of the main organizers of Latvian Olympiad in Informatics, was deputy or team leader of Latvian IOI and Baltic OI teams. M.Opmanis was head of jury of Baltic Olympiad in Informatics at BOI 1996, 1999, 2004 and 2012. From 2012 till 2015 he was a member of IOI International Committee.